Parameter clustering in Bayesian functional PCA of neuroscientific data

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

Parameter clustering in Bayesian functional PCA of neuroscientific data

Nicolò Margaritella* | Vanda Inácio | Ruth King

School of Mathematics, University of Edinburgh, Edinburgh, UK

Correspondence
*Nicolò Margaritella,
University of Edinburgh,
School of Mathematics,
James Clerk Maxwell Building,
The King's Buildings,
Peter Guthrie Tait Road,
Edinburgh, UK.
Email: N.Margaritella@sms.ed.ac.uk

Summary

The extraordinary advancements in neuroscientific technology for brain recordings over the last decades have led to increasingly complex spatio-temporal datasets. To reduce oversimplifications, new models have been developed to be able to identify meaningful patterns and new insights within a highly demanding data environment. To this extent, we propose a new model called parameter clustering functional Principal Component Analysis (PCl-fPCA) that merges ideas from Functional Data Analysis and Bayesian nonparametrics to obtain a flexible and computationally feasible signal reconstruction and exploration of spatio-temporal neuroscientific data. In particular, we use a Dirichlet process Gaussian mixture model to cluster functional principal component scores within the standard Bayesian functional PCA framework. This approach captures the spatial dependence structure among smoothed time series (curves) and its interaction with the time domain without imposing a prior spatial structure on the data. Moreover, by moving the mixture from data to functional principal component scores, we obtain a more general clustering procedure, thus allowing a higher level of intricate insight and understanding of the data. We present results from a simulation study showing improvements in curve and correlation reconstruction compared with different Bayesian and frequentist fPCA models and we apply our method to functional Magnetic Resonance Imaging and Electroencephalogram data analyses providing a rich exploration of the spatio-temporal dependence in brain time series.

KEYWORDS:
Bayesian hierarchical models, Clustering, Dirichlet process, Functional data analysis, Neuroscience, Spatio-temporal data

1 | INTRODUCTION

Several tools for the recording of different brain processes, such as functional Magnetic Resonance Imaging (fMRI) and Electroencephalogram (EEG) produce remarkable amounts of spatio-temporal data which challenge researchers to find suitable models for increasingly complex datasets. Consequently, the last decade has seen a marked increase in the development flexible methods for high dimensional data in neuroscience. Functional Data Analysis (FDA) is a fairly recent research field in statistics concerned with the analysis of data providing information about curves, shapes and images which vary over a continuum, usually time or space (see Ramsay and Silverman[1] for an overview). In the FDA framework, data can be considered as noise-corrupted,

discretised realisations of underlying smooth functions (curves or trajectories) which are recovered using basis expansions and smoothing.[2] Many standard statistical tools have been translated into the FDA framework. Functional Principal Component Analysis (fPCA) is a technique that defines a set of smooth trajectories as an expansion of orthonormal bases (eigenfunctions) and weights which are called functional principal component scores (fPC scores).[1] One of the advantages of fPCA is that it can be conveniently represented as a hierarchical mixed model in the Bayesian setting, with the joint posterior distribution of the fPC scores being the main target of inference.[3]

There has been a growing interest in applying FDA to neuroscientific data (see, among others, Viviani et al.,[4] Tian et al.[5] and Hasenstab et al.[6]). Often, in the FDA literature, underlying random curves are assumed to be independent and their correlation is ignored if believed to be mild.[7] However, curve dependence is of particular importance in the analysis of brain activity because of the complex architecture of spatio-temporal connections between brain areas.[8] Recently, Liu et al.[7] considered spatial dependence among trajectories by modelling the covariance of the fPC scores within a frequentist approach. Their results showed significant improvements in curve reconstruction compared to the standard approach assuming independence, especially with low signal-to-noise ratios.

The present study introduces a new method for the analysis of functional data in neuroscience. We develop a novel Bayesian fPCA model called Parameter Clustering fPCA (PCl-fPCA) that makes use of a Dirichlet Process (DP) mixture[9–11] to model the prior distribution of the fPC scores. Different functional mixture models that cluster functions through clustering of the coefficients in a basis expansion have been proposed in the literature.[12–19] However, these works have focused on a global clustering of curves, without considering local differences as well as the possibility of a dynamic evolution of dependence among curves. In this work we use the principal component bases due to their straightforward interpretation and employ DP mixture priors for every eigendimension retained. By allowing different clustering of the fPC scores for each eigendimension retained, we avoid the limitations of assuming separability of the cross-covariance and any a priori spatial covariance structure of the data, obtaining further insights from space-time interactions.

The study of how interactions among brain regions change dynamically during an experiment (i.e. dynamic functional connectivity) has recently attracted wide interest in the neuroimaging literature. This analysis has the potential to improve our understanding of how the brain works under both physiological and pathological conditions with recent studies focusing on the application of dynamic functional connectivity to aging,[20] schizophrenia,[21] dementia and Parkinson's disease.[22] This is a new frontier for neuroscientific research and the development of suitable models able to capture the intricate spatio-temporal dynamics in the data will lay the foundations for the progress in this area in coming years.[23]

In this regard, we show that our approach has multiple advantages in the analysis of neuroscientific data as it offers further insights into the spatio-temporal structure of the data as a result of dimension-specific curve classification; it improves curve reconstruction thanks to the local borrowing of information compared to current fPCA approaches; and it can be defined as a simple and computationally feasible hierarchical model which can be easily implemented in R.

The rest of the paper is structured as follows: in Section 2 we overview the standard Bayesian fPCA model and introduce our new method, along with computational details. Section 3 reports the setting and results of a simulation study where we compare the performance of PCl-fPCA with standard Bayesian and frequentist fPCA approaches under different data generating processes and noise levels. Section 4 addresses the application of our method to a resting-state fMRI dataset and a task-based EEG recording and we discuss the further insights obtained in the spatio-temporal structure of the data and the underlying neurophysiological processes. Conclusions are discussed in Section 5.

2 | METHODS

2.1 | Bayesian Functional PCA

The standard FDA model is given by

$$Y_{it} = X_{it} + \epsilon_{it}, \tag{1}$$

where Y_{it} denote the noise-corrupted, discretised, observed data for every spatially-correlated region (trajectory) $i = 1, \dots, n$ and time point $t = 1, \dots, T$; X_{it} the associated underlying random curve as a realisation of an L^2 stochastic process $\left\{X_t : t \in [1, T] \subseteq \mathcal{R}\right\}$ with mean μ_t and covariance function $G(s, t)$; and ϵ_{it} the noise term with zero mean and precision τ.[24]

Functional PCA assumes that the covariance kernel $G(s,t)$ of the process X_t can be represented by the Karhunen-Loève expansion, such that

$$G(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_{kt} \phi_{ks}, \quad s,t \in [1,T], \tag{2}$$

$$X_{it} = \mu_t + \sum_{k=1}^{\infty} \xi_{ik} \phi_{kt}, \quad i = 1, \dots, n, \tag{3}$$

where ϕ_{kt} are orthonormal eigenfunctions and λ_k are the associated eigenvalues. Then, each realisation X_{it} can be represented by a linear combination of eigenfunctions ϕ_{kt}, which are usually assumed to be observed, and fPC scores ξ_{ik}, which are the main goal of inference. The reader is referred to Chapter 8 of Ramsay and Silverman[1] and the recent review of Joliffe and Cadima[25] for a more detailed presentation of functional PCA. Although the number of eigendimensions can also be modelled with an appropriate distribution (see, for example, Suarez et al.[26]), this considerably increases the computational complexity of the model and thus in practice only K pre-determined terms of the linear expansion are retained pertaining to those that explain a sufficiently large part of the total variability in the data.[27] Often the case $\mu_t = 0$ is assumed and the centred data $\tilde{Y}_{it}$ are obtained by subtracting an estimate $\widehat{\mu}_t$ of the population average.[3]

The fPC scores ξ_{ik} are given prior probability distributions in the Bayesian framework. The standard Bayesian fPCA model[3] assumes fPC scores to be independent draws from a univariate zero-centred normal distribution whose variance is dependent on the eigendimension k. The most straightforward hierarchical representation of the standard Bayesian fPCA model is

$$\begin{aligned} \tilde{Y}_{it} &= \sum_{k=1}^{K} \xi_{ik} \phi_{kt} + \epsilon_{it}, \\ \xi_{ik} | s_k &\sim \mathrm{N}(0, s_k^{-1}), \\ \epsilon_{it} | \tau &\sim \mathrm{N}(0, \tau^{-1}), \\ s_k &\sim \Gamma(a,b), \\ \tau &\sim \Gamma(a',b'), \end{aligned} \tag{4}$$

with a, a', b, b' usually set to low values (e.g. 10^{-3}). In this model the noise term is assumed to be Gaussian and independent gamma priors are placed over the precision parameters because of their conjugacy property, permitting closed-form conditional posterior distributions and the use of Gibbs sampling.

Recently, Liu et al.[7] proposed to capture spatial dependence through a suitable model for the covariance of fPC scores. In particular, they defined $\mathrm{Cov}(\xi_{ik}, \xi_{i'k})$ as a function of the correlation coefficient $\rho_{ii'k}$ which they modelled using the Matérn function family and estimated the corresponding parameters. This approach implies the a priori definition of a covariance structure which depends on the distance between observations; such assumptions might not be suitable for complex spatio-temporal phenomena such as brain activity where dependencies are the result of both structural and functional neuronal pathways as well as task-specific characteristics. In this study, we overcome these limitations to achieve a higher level of flexibility in the modelling of the spatio-temporal covariance of neuroscientific data.

2.2 | PCl-fPCA model

In this section we present the structure of the PCl-fPCA model and the features of this approach that improve the current methods for functional PCA. The following hierarchical model defines the probability distribution generating observed time series. We present and comment on each level separately.

Level 1: As the standard Bayesian fPCA model in Equation (4), the distribution of the centred data given the parameters of the underlying smooth function and the noise term is given by:

$$\begin{aligned} \tilde{\mathbf{Y}}_i | \mathbf{X}_i, \tau &\sim \mathrm{N_T}(\mathbf{X}_i, \tau^{-1}\mathbf{I}), \\ \mathbf{X}_i &= \sum_{k=1}^{K} \xi_{ik} \boldsymbol{\phi}_k, \end{aligned} \tag{5}$$

where $\tilde{\mathbf{Y}}_i$, $\mathbf{X}_i$ and $\boldsymbol{\phi}_k$ are T-dimensional vectors and $\mathrm{N_T}(\mathbf{X}_i, \tau^{-1}\mathbf{I})$ denotes a multivariate Gaussian distribution with mean $\mathbf{X}_i$ and variance-covariance matrix $\tau^{-1}\mathbf{I}$ such that $\mathbf{I}$ denotes the $T \times T$ identity matrix. As in Equation (4), the eigenfuctions $\boldsymbol{\phi}_k$ are

assumed to be observed and the parameter τ does not depend on i or t, i.e. the noise is assumed to be constant in both space and time, although other characterisations are possible.[24] It follows that the likelihood function is given by

$$\mathrm{L}(\tilde{\mathbf{Y}}|\mathbf{X},\tau) = \left(\frac{\tau}{2\pi}\right)^{Tn/2} \exp\left\{ -\frac{\tau}{2}\sum_{i=1}^{n}(\tilde{\mathbf{Y}}_i - \mathbf{X}_i)'(\tilde{\mathbf{Y}}_i - \mathbf{X}_i)\right\}. \tag{6}$$

Level 2: To encode fPC scores cluster membership we introduce a classification variable c_{ik} as a stochastic indicator that identifies which latent class j in eigendimension k is associated with parameter ξ_{ik}. Prior distributions of the fPC scores ξ_{ik}, given the parameters of underlying clusters $\left[(\mu_{1k}, s_{1k}), \dots, (\mu_{Jk}, s_{Jk})\right]$ and the classification variable c_{ik}, are given by

$$\xi_{ik}|c_{ik}, \mu_{1k}, \dots, \mu_{Jk}, s_{1k}, \dots, s_{Jk} \sim \mathrm{N}\left(\mu_{c_{ik}}, s_{c_{ik}}^{-1}\right), \tag{7}$$

where $\mu_{c_{ik}=j}$ and $s_{c_{ik}=j}$ denote the mean and precision for the j-th cluster in the k-th eigendimension, respectively. Here we use a $J-$dimensional mixture of Gaussian distributions, independently, for each retained eigendimension $k = 1, \dots, K$ as we permit different (independent) partitions of the fPC scores for each mode of variation. It is worth recalling that, in the context of DP mixtures, J represents an upper-bound on the number of fPC score clusters.[28] In the rest of the manuscript we define $J_k^+ < J$ as the (data-driven) number of non-empty clusters in each eigendimension k.[29]

Level 3: Prior distributions for $\left[(\mu_{1k}, s_{1k}), \dots, (\mu_{Jk}, s_{Jk})\right]$ and $(c_{1k}, \dots, c_{nk})$, given hyperparameters r_k, β_k and parameters $(p_{1k}, \dots, p_{Jk})$, are given by

$$\begin{aligned} c_{1k}, \dots, c_{nk}|p_{1k}, \dots, p_{Jk} &\sim f_{\mathrm{C}}\left(p_{1k}, \dots, p_{Jk}\right), \\ \mu_{jk}|r &\sim \mathrm{N}(0, r_k^{-1}), \\ s_{jk}|\beta &\sim \Gamma(1, \beta_k), \end{aligned} \tag{8}$$

where f_{C} denotes the categorical distribution which generalises the Bernoulli random variable to J outcomes. Cluster precision s_{jk} can also be modelled using Uniform distributions on the cluster standard deviation where $\sigma_{jk} = 1/\sqrt{(s_{jk})}$.[30] Hyperparameters r and β are often centred around empirical estimates in the literature[31]; here, we take advantage of the properties of fPCA decomposition to tune the higher hierarchical levels in our model around weakly informative prior distributions. It follows from the Karhunen-Loève representation that, for any given i, ξ_{ik} are uncorrelated fPC scores with monotonically decreasing variance given by the eigenvalues λ_k[7]; therefore, sensible functions of the empirical estimates of the eigenvalues $\widehat{\lambda}_k$ can be used to fix r and β under the assumption that, for every eigendimension k, the position and dispersion of a cluster are both functions of $\widehat{\lambda}_k$. We note that setting $r = 1/\widehat{\lambda}_k$ and $\beta = \widehat{\lambda}_k$ worked well in our simulations and application.

Level 4 and 5: Prior distribution for $(p_{1k}, \dots, p_{Jk})$, given hyperparameter α and prior distribution for α are given by

$$\begin{aligned} p_{jk}'|\alpha_k &\sim \mathrm{Beta}(1, \alpha_k), \\ p_{1k} = \frac{p_{1k}'}{\sum_{j=1}^{J} p_{jk}'}; \quad p_{jk} &= \frac{p_{jk}' \prod_{l<j}(1 - p_{lk})}{\sum_{j=1}^{J} p_{jk}'}, \quad j = 1, \dots, J \\ \alpha_k &\sim \mathrm{U}[0, Q_k], \end{aligned} \tag{9}$$

where p_{jk} follow the stick-breaking construction[32] with parameter α_k modelling the prior belief over the mixing proportions $p_{1k}, \dots, p_{Jk}$. The dispersion parameter α is usually fixed or modelled with a prior distribution; here we used a uniform distribution with sufficiently large Q.[11,33–35]

Different specifications of s_{jk} and Q can be employed for $k = 1$ and $k = 2, \dots, K$ to incorporate the knowledge that the first eigendimension is more likely to capture global patterns in the data while the following dimensions are more sensitive to local features. For example, in the first eigendimension one can use the gamma distribution for the cluster precision in Equation (8) as it assigns more weights to large clusters than a uniform on the standard deviation which can be used instead in the subsequent dimensions. We provide specific examples in Section 3.1 and the results of a sensitivity analysis on Q, β and s in the WebA section of the Supplementary Material file.

The model structure can be displayed with a direct acyclic graph (DAG) (Supplementary Material, WebB section, Figure 1). As J approaches infinity the model corresponds to a DP mixture model[10,11,33,34,36] with the difference that we have placed here multiple independent mixtures over the prior distribution of the fPC scores. In practice we used the truncated stick-breaking construction and tested the model with different commonly chosen values of J ($J = 20, 30$ and 50). The upper bound J should be chosen sufficiently large to ensure $J_k^+ < J$ in each eigendimension. Larger Js will naturally impact on computations (e.g.

in our applications we observed the computational time of the model with $J = 50$ to be ~ 1.5 higher than with $J = 20$). All the conditional posteriors of this model (most of them available in closed form) are provided in Appendix A. Markov chain Monte Carlo (MCMC) techniques are used to simulate from the joint posterior distribution of all parameters given the data. Reconstruction of the smooth trajectories x_{it} is made easy by its linear relationship with the model parameters ξ_{ik}; thus it is possible to obtain the posterior distribution of the i-th curve for every t and at every MCMC iteration w,

$$x_{it}^{(w)} = \bar{x}_t + \sum_{k=1}^{K} \xi_{ik}^{(w)} \phi_{kt}, \;\; i = 1, \dots, n; \;\; t = 1, \dots, T, \;\; w = 1, \dots, W, \tag{10}$$

where $\bar{x}_t$ is the smoothed estimate of the sample mean $\sum_{i=1}^{n} y_{it}/n$. It follows that symmetric 95% point-wise credible intervals for each trajectory-specific mean can be obtained easily from Equation (10) by considering the $(1-\alpha)/2$ and $\alpha/2$ quantiles of the $\left\{ x_{it}^{(1)}, \dots, x_{it}^{(W)} \right\}$ empirical distribution.

2.3 | Clustering

In this section we focus on the clustering of fPC scores. The discrete nature of the DP is very useful for clustering as it allows ties among the latent c_{ik} [37]; therefore, DP mixtures implicitly return classification through the allocation of each fPC score to a generating distribution with some probability. Clustering uncertainty can be evaluated at different levels such as the number of clusters, the size of each cluster and the fPC scores assigned to them. For the explorative purpose of our model we avoid the use of automated algorithms to select a final partition of the fPC scores (either classical hierarchical or partitioning algorithms based on the similarity matrix [34] or more recently proposed algorithms based on a loss function over clusterings [38]). Instead, we propose a 3-step exploration of the empirical distribution of generated clusterings which we find useful to evaluate clusters uncertainty arising from the data. After burn-in, the empirical distribution of generated clusterings $\left\{ \boldsymbol{c}_k^{(1)}, \dots, \boldsymbol{c}_k^{(W)} \right\}$ can be considered a good approximation of the true posterior distribution [10] and it can be used to obtain other distributions of interest, such as the number and size of non-empty clusters, maximum a posteriori probabilities (MAPs) and pairwise probability matrices (PPMs). We make use of these distributions in a 3-step exploration.

Step 1: The distribution of the number of non-empty clusters J_k^+ can be obtained by exploring the values of the classification variable $\boldsymbol{c}_k$ for all the W iterations retained after burn-in $\left(J_k^{+,w} = \max_j \left\{ \boldsymbol{c}_k^w \right\} \right)$. Although considering the number of non-empty clusters J_k^+ does not account for size and stability (i.e. the number of times a cluster appears in the MCMC chain), the distribution of J_k^+ provides a useful first check for assessing the presence of more than one cluster in each eigendimension. For this purpose, we used the Bayes Factor (BF) defined as $\left\{ P_\pi(J_k^+ = 1)/P_\pi(J_k^+ > 1) \right\} \times \left\{ P(J_k^+ > 1)/P(J_k^+ = 1) \right\}$ where $P_\pi(J_k^+ = j)$ denote posterior probabilities and $P(J_k^+ = j)$ the relative prior probabilities which can be obtained by simulating from the prior distribution of $\boldsymbol{c}_k$. A BF greater than 1 suggests absence of clusters in the fPC scores of a specific eigendimension; hence, this step identifies those eigendimensions where clusters are more likely to exist in the data.

Step 2: The distribution of the cluster size can be obtained by counting for each iteration w the number of fPC scores allocated to the same label $\left(\sum_{i=1}^{n} \mathrm{I}\left(c_{ik}^{(w)} = j \right), \forall j \in \left[1, J_k^+\right] \right)$ or by monitoring the posterior distribution of the mixing proportions p_{jk}. Although there is no guarantee that fPC scores joining a cluster remain loyal to it, the size of clusters permits the identification of clusters which are populated only sporadically as a result of the uncertainty in the classification of subsets of fPC scores. The distribution of these clusters has typically a notable probability mass at zero. Therefore, this second step can help understand the number and dimension of clusters we expect to see in each eigendimension and the relative uncertainty.

Step 3: Finally, MAPs and PPMs can help refine our understanding of the underlying clustering. MAPs are commonly used to identify the most probable clustering for each observation and they can be computed by identifying for each fPC score the posterior mode of c_{ik} from the empirical distribution of generated clusterings. MAPs are known to be limited by the possible presence of multiple modes and cases where individuals who share the same modal group are less frequently together than with others in different clusters. These issues can be addressed by the PPMs which represent the posterior belief for all pairs of curves to belong to the same cluster regardless of the clustering label. [33,34,36] For each iteration w, an $n \times n$ association matrix $\delta(\boldsymbol{c_k})$ can be obtained with indicators $\delta_{ii'}(\boldsymbol{c_k})$ which takes value 1 if fPC score i and i' in eigendimension k are clustered together and 0 otherwise. Element-wise averaging over all these association matrices yields the PPM. Combining the exploration of MAP and pairwise probabilities can narrow down a decision on the most likely partition of the fPC scores.

Although we find limitations for each of these steps individually to draw robust conclusions, considering them together as a whole provides rich information on the (a posteriori) most likely partition for each eigendimension. Particularly in the case of complex phenomena, such as those captured by neuroscientific recordings, a thorough exploration of cluster uncertainty in the

data should be always considered to ensure a sensible interpretation of the results. We present an application of these analyses to fMRI and EEG data in Section 4. In a Bayesian mixture model where cluster identification is of interest, extra care should be taken to avoid label switching arising from the symmetry in the likelihood of model parameters. This can be avoided either by imposing identifiability constraints on the parameter space or by employing relabelling algorithms. In our simulation study and applications we found that imposing constraints on the order of cluster means ($\mu_{1k} <, \dots, < \mu_{Jk}$) or weights ($p_{1k} <, \dots, < p_{Jk}$) was enough to successfully control label switching.

2.4 | fPC score clustering as generalisation of standard clustering

In the standard infinite mixture model based clustering, the indicators $c_i = c_{i'} = j$ with $i \neq i'$ would associate a couple of trajectories to a certain cluster j with probability $P_{ii'}$. On the other hand, by placing infinite mixtures over the fPC scores for every eigendimension retained, we allow for a more complex network of dependence among curves. In our model, c_{ik} and $c_{i'k}$ would associate fPC scores i and i' to potentially different clusters in every eigendimension k with probability $P_{ii'k}$. It follows that a pair of curves could happen to share the same cluster in only part of the K eigendimension retained, expanding the standard model based clustering to a richer classification method. Furthermore, as each dimension represents a mode of variation (eigenfunction) and its importance (eigenvalue), our method offers additional insights into the underlying spatio-temporal structure of the data. In the following sections we show how clustering fPC scores produces a rich spatio-temporal exploration of complex neuroscientific data.

3 | SIMULATION STUDY

3.1 | Simulation scenarios

We performed a simulation study to assess the performance of PCl-fPCA model and compare it to the standard Bayesian fPCA model in terms of both curve reconstruction and classification for different data generating processes and noise levels. We also included for comparison two frequentist approaches: the standard fPCA model[1] and a modified version of the model by Liu et al.[7] that we adapted to the features of neuroscientific data. In this latter model, curve dependence is captured through the fPC scores by means of independent Matérn functions for each eigendimension retained.

In order to test model performance with simulated data matching those of the targeted neuroscientific applications as closely as possible, we generated two eigenfunctions from simulated data resembling evoked responses in the brain using the function `pca.fd` from the `fda` package in `R`[39]. Subsequently, we defined three data generating processes (DGP) that differ in the way the fPC scores are generated: in the first DGP (DGP1), scores are generated from different mixtures of Gaussian distributions in the two eigendimensions considered; in the second DGP (DGP2), fPC scores dependence in the first eigendimension is generated from a Matérn function while in the third DGP (DGP3), dependence of fPC scores is generated by independent Matérn covariance functions with different parameter values in each eigendimension. For each DGP, we combined the two eigenfunctions with the fPC scores to build the simulated datasets. We applied a random Gaussian noise and tested the models with both high and low signal-to-noise ratios (STN=6 and 1 respectively). Figure 1 shows an example from the set of 100 generated curves in DGP1 where either a low or high random noise is added.

One hundred datasets ($L = 100$) for each DGP and STN were input to fPCA first for curve smoothing using cubic B-splines and dimension reduction by estimating the respective eigenvalues and eigenfunctions using the function `pca.fd` from the `fda` package in `R`[39]. We retained a number of dimensions K explaining at least 95% of the total variability in curves. Figure 1 shows eigenfunctions and their weights extracted after smoothing a set of low-noise curves for the first DGP.

We adapted the general model presented in Section 2.2 to the specific simulation analysis using eigenvalues λ_k and their properties to develop vaguely informative prior distributions for the parameters r, β and Q (Equations (8) and (9)) in the two eigendimensions retained $k = 1, 2$. We set $r \in \left\{1/\hat{\lambda}_1,\ 1/\hat{\lambda}_2\right\}$ and $Q \in \{10, 5\}$ as well as setting $s_{j,1} \sim \Gamma(1, \lambda_1)$ and $\sigma_{j2} \sim \mathrm{U}[0, \sqrt{\lambda_2}]$. The use of a uniform distribution in the second dimension favours the search of smaller clusters than in the first eigendimension, as increasingly local features should be expected in trailing modes of variation.[7] We made sure that even the smallest upper-bound Q of the dispersion parameter α distribution represented an expected number of clusters a priori far higher than the ground truth.[40,41] A similar choice for α was specified by De Iorio et al.[35] due to the resulting stable computations.

We coded the model in `R` using the `rjags` package[42], and employed a conservative approach using 100, 000 iterations for the burn-in and retaining the subsequent 100, 000 MCMC iterations.[33,43] The convergence diagnostics did not suggest lack of

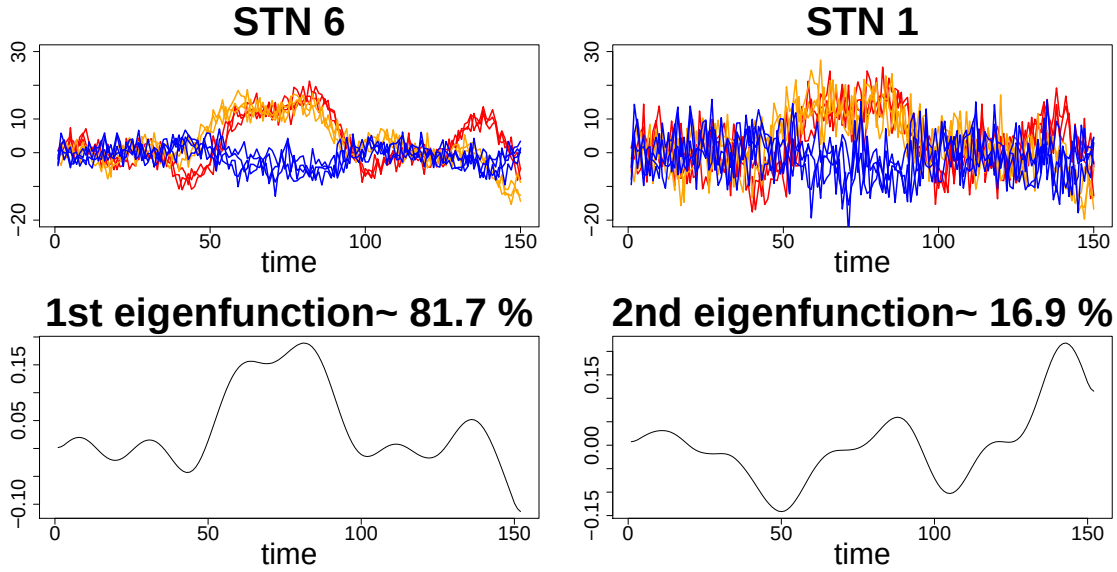

FIGURE 1 Simulation study: (Top) an example of curves from DGP1 with low random noise (STN6) and high random noise (STN1). (Bottom) the first and second eigenfunctions extracted from a set of DGP1 curves with STN6. This figure appears in colour in the electronic version of this article.

convergence for all the parameters of interest. We used a thinning of 5 to store results from 100 simulated datasets efficiently (approximately 70 MB each with $K = 2$). It takes 36 minutes on average to complete one simulation run on a 2-core Intel CPU running at 2.7 GHz with 8 GB RAM.

We used Integrated Mean Squared Error (IMSE) to measure and compare reconstruction performance between PCl-fPCA model and the competitor models. IMSE and its associated approximation for every curve i are given by

$$\text{IMSE}_i = \text{E}\left\{ \int \left(\hat{x}_{it} - x_{it}\right)^2 \text{dt} \right\} \approx \frac{1}{L} \sum_{l=1}^{L} \left\{ \frac{1}{T} \sum_{t=1}^{T} \left(\hat{x}_{i_l t} - x_{it}\right)^2 \right\}, \tag{11}$$

where the expectation is taken with respect to the underlying curve x_i. The IMSE is a useful measure of performance in density estimation and is frequently used in curve reconstruction.[44,45] In addition, as curves correlation $\rho_{ii'}$ is often of interest in neuroscientific applications (e.g. for measuring the degree of functional connectivity between brain areas), we measured correlations reconstruction using the L2 norm $||\widehat{\rho}_{ii'} - \rho_{ii'}||_2$ and compared it with those of the competitor models.

In order to assess the proposed model clustering performance in DGP1, we adopted the Adjusted Rand Index (ARI) to quantify the similarity between the estimated partitions (using MAP) and the ground truth for every simulated dataset l and eigendimension k. The ARI is commonly used in the literature to assess clustering performance as it varies between exact partition agreement (1) and when partitions agree no more than is expected by chance (0).[36,46] Moreover, we measured the improvement in distance (L_2 norm) between the posterior pair-wise probability matrices and the ground truth to evaluate the clustering performance of PCl-fPCA model by taking into account cluster uncertainty. Further details on the simulations setting can be found in WebC section of the Supplementary Material.

3.2 | Simulation results

Results of curve and correlation reconstruction are reported in Figure 2. The case where STN = 1 is particularly relevant because neuroscientific data are usually affected by high noise. In this scenario, PCl-fPCA model highly improved curve reconstruction compared to all competitor models as 100% of the true curves were better recovered under PCl-fPCA and the median improvement in IMSE ranged from 22% to 45%. Moreover, a similar improvement was also obtained for DGP2 where clustering is present in only one eigendimension (Figure 2, bottom left). In addition, correlation reconstruction was also better achieved under PCl-fPCA with a median percentage of improvement ranging from 20% to 30% for DGP1 and 2% to 8% for DGP2 (Figure 2, right column). In the case of low noise (STN6), the proposed model still performed better than the competitors for DGP1 and

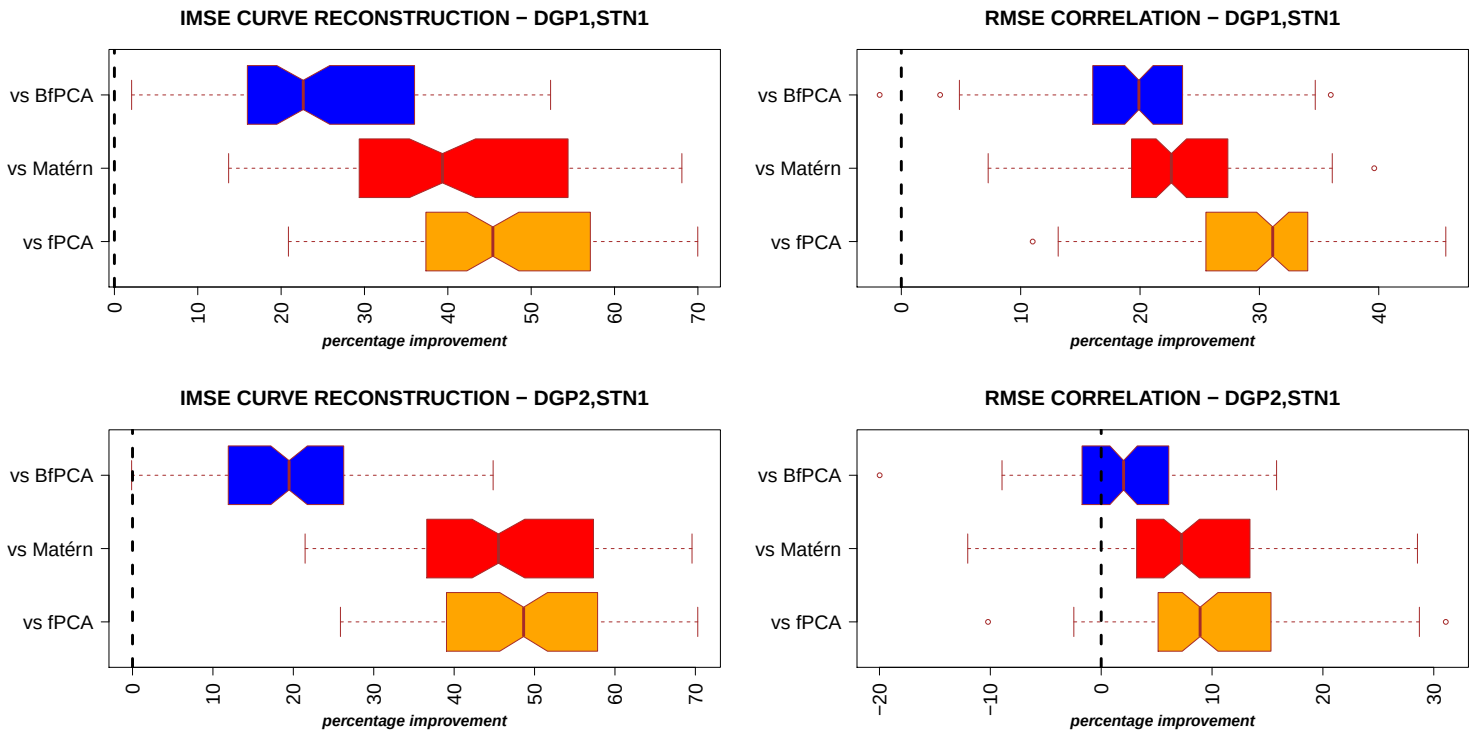

FIGURE 2 Simulation study: curve and correlation reconstruction for Data Generating Processes (DGP) 1 and 2 with high noise (STN1). IMSE and RMSE improvement percentage using PCl-fPCA model versus standard Bayesian fPCA (BfPCA), fPCA model for correlated curves (Matérn) and standard fPCA model (fPCA). This figure appears in colour in the electronic version of this article.

TABLE 1 Simulation study: clustering performance of PCl-fPCA in DGP1. The table reports median and interquartile range of ARI computed for each simulated dataset and every STN and eigendimension analysed.

Eigendimension	*ARI*
STN=1	
1st dim	1 [1,1]
2nd dim	0.753 [0.444,0.868]
STN=6	
1st dim	1 [1,1]
2nd dim	0.966 [0.933,0.966]

achieved values of IMSE and RMSE similar to those of the best competitor models in DGP2 (Supplementary Material, WebB, Figure 2). Interestingly, even when no clusters are expected in both eigendimensions (DGP3), the performance of the PCl-fPCA was still comparable to the best ones achieved by competitor models for both low and high noise levels (WebB, Figure 3).

The performance of the PCl-fPCA model in terms of classification is reported in Table 1. The proposed model scored high in the ARI classification index in both eigendimensions studied; two and three clusters were expected in the first and second dimension respectively in DGP1. Clusters in the first eigendimension were always correctly identified by ARI for both high and low signal to noise ratios. The identification of three clusters in the second eigendimension was more challenging as they were smaller and nearer to each other; however, scores near 1 were almost always obtained when the low noise scenario was tested and even in the case of high noise we observed fairly high scores. Similar results were achieved by measuring the improvement in

distance (L2 norm) between the posterior pair-wise probability matrices and the ground truth to account for cluster uncertainty in the classification performance (WebC, Table 2).

Figure 4 in section WebB of the Supplementary Material provides evidence of the improved level of information achieved by PCl-fPCA in the DGP1 scenario. Overall, PCl-fPCA model outperformed the competitors in curve reconstruction under different data generating processes, especially in the case of high noise in the data; moreover, for the case where clusters are not limited to one eigendimension, the proposed model was able to retrieve the original spatial partition in each eigendimension and bring to light important relationships between clusters. These results could further help the understanding of underlying neuroscientific phenomena in a real data scenario.

4 | APPLICATION

In this section we present two applications of the PCl-fPCA model to the analysis of neuroscientific data from fMRI and EEG recordings. In Sections 4.1 and 4.2, the PCl-fPCA model is used to explore underlying brain patterns arising from a short time window fMRI recording of a healthy subject at rest. In the emerging field of dynamic functional connectivity, the analysis of the evolution of brain patterns within a short time window is of particular interest as it could uncover transient configurations of coordinated brain activity.[47] The aim of the present fMRI analysis is to verify whether the results obtained on a short time window recording (1 minute) are in line with the current knowledge on brain resting-state networks obtained from static functional connectivity studies where results are typically averaged over 5-15 minutes recordings. In Sections 4.3 and 4.4 the PCl-fPCA model is used for artefacts identification in the EEG recording of a healthy subject under a two-stimuli paradigm (match vs unmatch images). The presence of artefacts originating from sources different from the brain and contaminating brain signals is a well-known problem in EEG recordings and an active area of research in neurophysiology.[48] The aim of the present EEG analysis is to check whether the fPC-PCA model can be successfully used to identify the spatio-temporal features of different artefacts and the location of the relative affected brain areas.

4.1 | fMRI setting

The study relates to a thirty-year-old healthy woman volunteer who underwent a resting-state fMRI at the Department of Radiology, Scientific Institute Santa Maria Nascente, Don Gnocchi Foundation (Milan, Italy) during February 2015. The recording was carried out using a 1.5 T Siemens Magnetom Avanto (Erlangen, Germany) MRI scanner with 8-channel head coil. The subject was asked to lie down in the MRI machine in supine position with eyes closed while Blood Oxygenation Level Dependent Echo Planar Imaging (BOLD EPI) images were acquired. She was instructed to keep alert and relaxed; no specific mental task was requested.

High resolution T1-weighted 3D scans were also collected to be employed as anatomical references for fMRI data analysis. Standard pre-processing involved the following steps: motion and EPI distortion corrections, non-brain tissues removal, high-pass temporal filtering (cut-off 0.01 Hz) and artefacts removal using the FMRIB ICA-based Xnoiseifier (FIX) toolbox.[49] After the pre-processing, the resulting 4D dataset was aligned to the subject's high-resolution T1-weighted image, registered to MNI152 standard space and resampled to $2\times2\times2\ mm^3$ resolution. One minute length series (sampled at 0.5 Hz) were extracted as the average signal within each of 90 regions of interest (ROIs) according to the Automated Anatomical Labeling (AAL90) coordinates. The resulting 30×90 dataset was input to fPCA for curve smoothing and dimension reduction using the `pca.fd` function from the `fda` package in `R`.[39] The set of 90 smooth curves and the retained eigendimensions are shown in Figure 5 of Supplementary Material WebB. We kept the first three dimensions explaining more than 85% of the total variability while accounting for more than 10% each.

We adapted the general model in Section 2.2 following the approach taken in the simulation study (Section 3.1), favouring global patterns in the first eigendimension and local patterns in the remaining dimensions. We assessed convergence using trace plots and BGR diagnostics and the number of independent retained samples by computing the effective sample size (WebD, supplementary material). We employed the same computational approach described in Section 3.1 and it took 59 minutes to run the analysis with $K = 3$ on a 2-core Intel CPU running at 2.7 GHz with 8 GB RAM. Furthermore, we carried out a sensitivity analysis by varying the values of the hyperparameters β, Q and the distribution of s in each dimension (WebA, Supplementary Material).

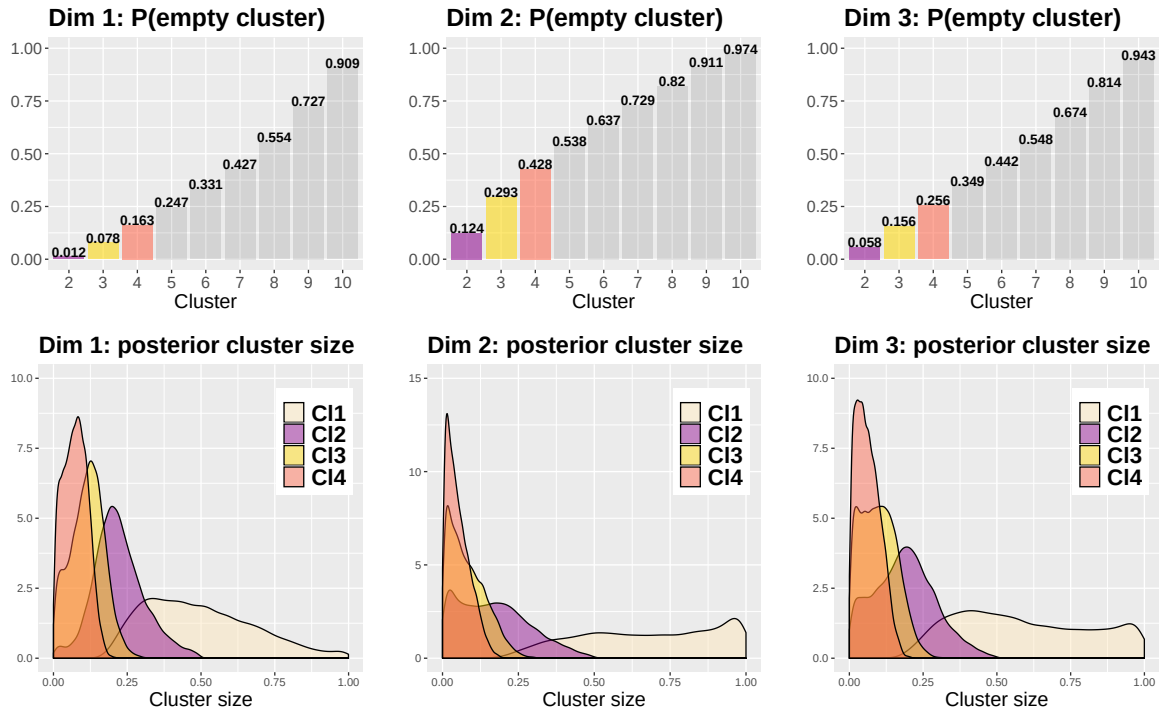

FIGURE 3 fMRI data analysis: cluster identification. The first row shows the posterior probabilities of being empty for the second to tenth clusters in the three eigendimensions (Dim 1:3) analysed. The second row shows the posterior distributions of cluster size (given it is not empty) among the first four clusters (Cl1:Cl4, right to left). This figure appears in colour in the electronic version of this article.

4.2 | fMRI analysis results

The posterior probabilities associated with the single cluster (i.e. no clusters) scenario were 0.012, 0.124 and 0.058 for the three eigendimensions k, respectively. The Bayes factors for the first eigendimension was 0.53, which indicates some evidence against no clusters. Conversely, the second and third dimensions returned BF = 2.93 and 1.33 respectively, which can be interpreted as evidence in favour of a single cluster. It is worth noting that, as the implied prior probabilities were highly in support of multiple clusters, the BF for $k = 2$ and 3 show a diametrical change from prior to posterior belief. These results are also confirmed by a BF sensitivity analysis which is reported in the supplementary material (WebA).

Figure 3 shows the posterior probability for a cluster being empty and the posterior distributions of cluster size given it is not empty. Two to three clusters seem to emerge in dimension 1; the size of the second cluster (Cl2, second from the right in Figure 3, bottom-left panel) has a peak around 20%, very small mass near zero, and a very low probability of being empty. The third cluster (Cl3) has a size peaking at 12% but more mass near zero and a higher probability of being empty. On the other hand, dimension 2 and 3 seem to suggest the presence of no more than one cluster each. The second cluster in both these dimensions has higher probability of being empty and the distributions of size have much more mass around zero. Furthermore, the distributions of the first cluster (Cl1) in both dimensions have a notable peak around 90% suggesting that, even when more than one cluster is considered, the large majority of fPC scores in dimension 2 and 3 tends to be gathered within a single large cluster.

The use of MAPs suggests there might be no more than 2 groups in the first dimension and 1 group in the second and third dimensions. Clustering with MAPs in the first dimension identified 9% of curves whose trajectories are wigglier and with a visibly shorter inter-peak difference between the first positive and negative peaks compared to the other group (WebB, Figure 6). Figure 7 of section WebB in the Supplementary Material shows an example of curve reconstruction using the posterior mean and 95% point-wise credible bands of the subject specific mean. Curves in cluster 2 pertain to brain areas from the occipital lobe (Calcarine, Cuneus, Lingual, Inferior Occipital Gyrus) and parietal lobe (Precuneus).

By analysing the pairwise probability matrix, a more comprehensive classification emerged. The previously dichotomous partition in dimension $k = 1$ is now enriched by a third group of brain areas with no clear clustering preference (grey band at the

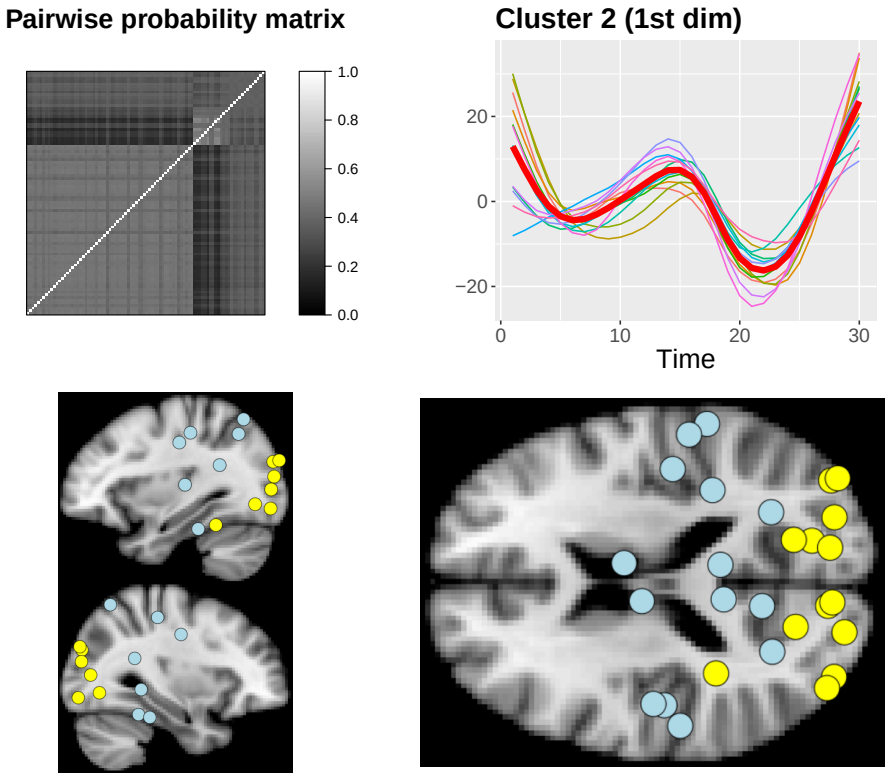

FIGURE 4 fMRI data analysis: cluster identification with pairwise probabilities. Top-left: pairwise probabilities suggesting a tripartition of curves in the first eigendimension. Top-right: cluster 2 updated according to the partition suggested by pairwise probabilities. The thick line represents the cluster mean. Bottom: the 3-D representation of clusters 2 and 3 over sagittal and axial slices of the human brain, where yellow (light) dots represent locations in cluster 2 and blue (dark) dots those in cluster 3. This figure appears in colour in the electronic version of this article.

top-right of the pariwise probability matrix in Figure 4). Cluster 2 comprises 16% of curves which all represent areas from the occipital lobe (yellow-light dots), while curves in cluster 3 (blue-dark dots) belong to the cingulate cortex (Middle and Posterior Cingulate Cortex), parietal (Parietal Superior Lobule, Precuneus) and temporal (Middle and Inferior Temporal Gyrus) lobes (Figure 4, a colour version of this figure can be found in the online version of the article).

We note that these three clusters are supported in the neuroimaging literature. It is well established that primary and extra-striate visual regions are active at rest[50] and have a role in processing mental imagery.[51] Just outside the visual cortex, the Temporal Inferior Gyrus takes part to the visual ventral stream which links information from the visual cortex to memory and recognition.[52] Moreover, the Posterior Cingulate Cortex is known to interact with several different brain networks simultaneously and it participates in the Default Mode Network together with part of the parietal lobe.[53] Conversely, it has been suggested that areas pertain to the Prefrontal Cortex (all included in cluster 1) have less long-range connectivity in the resting state condition.[54] Finally, the sensitivity analysis further confirmed our findings as they were robust to changes in both shape and value of the hyperparameters (WebA, Supplementary Material).

4.3 | EEG setting

For our second application we employed data from an EEG study on brain activations following object recognition tasks (Event Related Potentials, ERPs).[55] ERPs are very small bio-electrical signals generated by the brain in response to specific events or stimuli. They are EEG changes time locked to motor, sensory or cognitive events that provide a non-invasive approach to study

psychophysiological correlates of mental processes.[56] In contrast, body or eye movements introduce large artefacts to EEG recordings and trials contaminated with artefacts need to be corrected or even discarded.[57] In the present study we employed the PC-fPCA model for artefacts identification in the EEG recording of a single healthy subject. The individual was presented with two separate stimuli in the forms of images taken from the 1980 Snodgrass and Vanderwart picture set.[58] The second stimulus was either a different image (unmatch) or the same image (match) as in the first stimulus. We used the data-driven clustering of the PCl-fPCA model to identify the spatio-temporal features of different artefacts and the relative affected brain areas.

The data were recorded using a cap with 64 electrodes placed on the subject's scalp and the brain activity at each recording electrode was sampled at 256 Hz for 1 second. Further details on the recording setting can be found in Zhang et al.[55] We considered both the unmatched and matched tasks within the same analysis and used our PCl-fPCA model to find data-driven differences in the morphology of the curves. Therefore, a 128×256 dataset was input to fPCA for curve smoothing and dimension reduction using the `pca.fd` function from the `fda` package in `R`.[39] The set of 128 smooth curves and the retained eigendimensions are shown in Figure 8 of Supplementary Material WebB. We kept the first two dimensions explaining 90% of the total variability while accounting for more than 10% each. We applied the same model settings described in Section 4.1; we assessed convergence using trace plots and BGR diagnostics and the number of independent retained samples by computing the effective sample size (WebD, supplementary material). We employed the same computational approach described in Section 3.1 and it took 64 minutes to run the analysis with $K = 2$ on a 2-core Intel CPU running at 2.7 GHz with 8 GB RAM.

4.4 | EEG analysis results

Two clusters seem to emerge in dimension 1. The size of the second cluster (Cl2, second from the right in Figure 5, bottom-left panel) has a peak around 20%, and a low probability of being empty. The third and fourth clusters (Cl3, Cl4) have both sizes peaking near zero and higher probabilities of being empty. On the other hand, dimension 2 clearly indicates the presence of three clusters with sizes peaking at 60%, 20% and 20% and very low probabilites of being empty. Furthermore, the distributions of the first cluster (Cl1) in both dimensions have very low mass near 1, supporting the presence of multiple clusters in both dimensions.

Both MAP and pairwise probability analyses confirmed the presence of 2 clusters in the first dimension and 3 clusters in the second dimension (Figure 6). The second cluster in the first eigendimension contains all the recordings from electrodes in the frontal areas for both the matched and unmatched tasks (Figure 9, WebB, Supplementary Material). These curves have a marked peak at the end of the recording, indicating a possible artefact (probably originated from eye blinking), and they appear to have two separate underlying patterns. These trends are captured in the clustering of the second eigendimension where the second and third clusters further divided the EEG activity in the frontal brain areas between those recorded during the matched and unmatched tasks (Figure 9, WebB, Supplementary Material). Notably, despite all curves showing more variability toward the end of the recordings, we found that only those from frontal areas have a consistently different behaviour from that of the group. This is in line with the work of Zhang et al.[55], where the authors excluded frontal region recordings from part of their analyses because of an inconsistent wave morphology compared with the wave form of the other regions. Frontal areas are known to be prone to recording artefacts particularly from eye movement which might have affected the different wave forms observed in these data.[57] Furthermore, the data-driven separation of frontal area curves into tasks (matched and unmatched) suggests the effect of two separate artefacts on the amplitude of these recordings.

5 | DISCUSSION

The processing of the human brain is a complex phenomenon in both time and space. The modelling of spatio-temporal datasets in the big data era is a challenge becoming every day more demanding as we struggle to keep up with the overwhelmingly larger datasets we are required to make sense of. Moreover, the extraordinary advancements in neuroimaging of the last decades have focused large part of neuroscientists and statisticians' efforts on the spatial domain both in clinical practice and research (see, for example, Durante et al.[59]). Nevertheless, the study of how interactions among brain regions change dynamically during an experiment, (i.e. *dynamic functional connectivity*) has recently attracted interest in the neuroimaging literature.[60] In fact, the time domain retains important neurophysiological information on brain functioning and neuronal health and without it we are at risk of drawing partial and possibly wrong conclusions on how the brain works.

In the present study we proposed a model that combines functional PCA and Bayesian nonparametric techniques to explore spatio-temporal datasets flexibly. We combined the idea of introducing spatial dependence among curves through the fPC scores

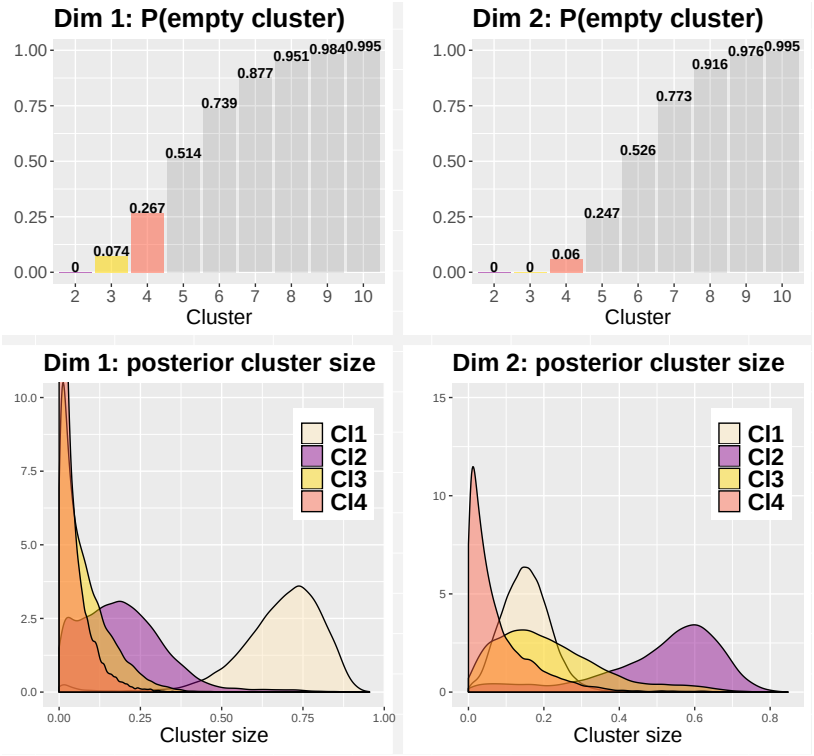

FIGURE 5 EEG data analysis: cluster identification. The first row shows the posterior probabilities of being empty for the second to tenth clusters in the two eigendimensions analysed. The second row shows the posterior distributions of cluster size (when not empty) among the first four clusters (Cl1:Cl4, right to left). This figure appears in colour in the electronic version of this article.

proposed by Liu et al.[7] with the infinite Gaussian mixture model to obtain a flexible modelling of the covariance structure. The main results show a clear improvement of the PCl-fPCA model both in curve and correlation reconstruction compared to different state-of-the-art fPCA models, particularly in the presence of high noise (as it is often the case in brain recordings) and the ability of exploring curves dependence dynamically allowing for different spatial patterns for each eigendimension retained.

Improvements in the reconstruction of high-noise corrupted curves were also reported by Liu et al.[7]; in fact, the beneficial effect of accounting for curves similarity is more evident when the true signal is well masked behind the noise. Nevertheless, a direct modelling of large covariance matrices often resorts to the use of common covariance functions to avoid overparametrisation. The use of functions such as Matérn or rational quadratic implies a priori knowledge on the shape of spatial dependence. We believe that this approach does not suit highly complex phenomena, such as brain processing, where dependence has a much more elaborate architecture than a simple function of spatial proximity. Clustering the fPC scores allowed us to capture dependence among curve flexibly without the need to estimate the relative spatial covariance matrix. Interestingly, our results suggest that the high flexibility of PCl-fPCA model makes it a very suitable choice even in the cases where a single or even none of the eigendimensions retained support clustering of fPC scores. Further improvements may be derived from modelling the correlation or autocorrelation structure of the noise, although the trade-off with model complexity should be taken into account.[1]

DP mixture models have also been used for clustering time series through the clustering of the relative coefficients in a basis expansion representation. Many of these works have focused on global clustering, where curves are clustered together *for all* their coefficients.[12–19] However, not only in neuroscientific data, but in many other types of functional data, curves might be characterised by regions of heterogeneous behaviours[61]; therefore, some authors have proposed alternative approaches that allow also for local differences in the clustering.[62,63] In the present study we moved from a global clustering of the data to a local clustering of fPC scores to address both the exploration of brain activity data and to improve curve reconstruction. Dunson et al.[62] and MacLehose et al.[63] used local clustering only as a means to improve estimation and their methods either neglect inter-subject variability in the coefficients (Dunson et al.[62]) or lack cluster interpretability (MacRose at al.[63]). In contrast, our approach combines the straightforward interpretation of the eigenfunctions with a local clustering of the fPC scores which

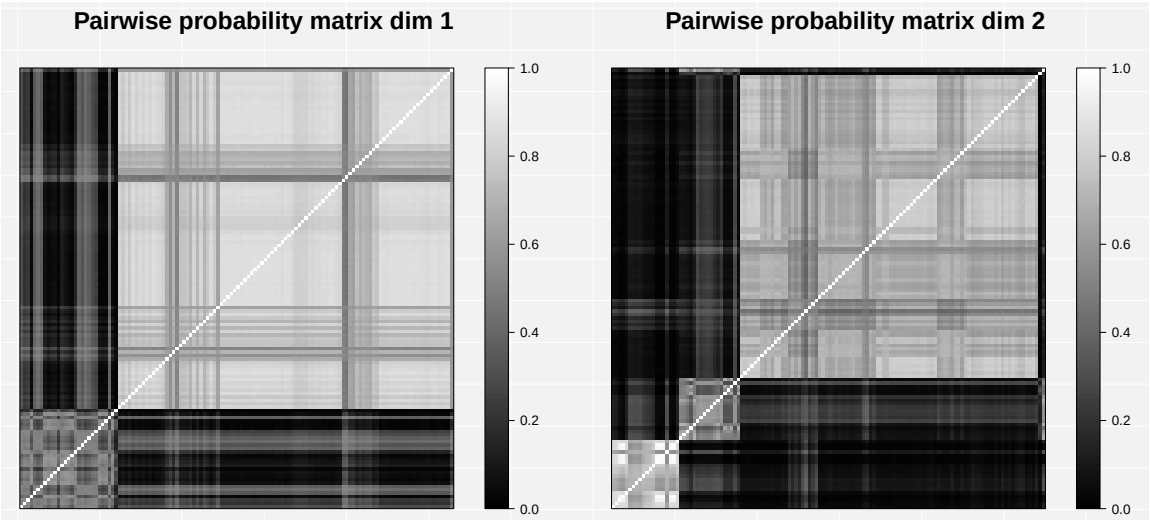

FIGURE 6 EEG data analysis: cluster identification with pairwise probabilities. First column: pairwise probabilities suggesting a bipartition of curves in the first eigendimension (top) and a tripartion in the second (bottom).

account for inter-subject variability within each cluster. Therefore, we obtained both an improved curve reconstruction and a rich classification technique. In fact, curves are never identical, they can be potentially assigned to different clusters in each eigendimension, and each eigendimension can have a different number of clusters (see Figure 4 of Supplementary Material WebB for a visual example). In addition, the assumption of separability of the cross-covariance matrix is avoided and complex time-space interactions are captured by the model; as a consequence, this local borrowing of information also improves the reconstruction of the underlying smooth process. Moreover, we benefit from the properties of the fPCA expansion to tune the hyperparameters and improve the MCMC convergence.

Cross-covariance matrices are often intractable if we do not resort to compromises in our models. A sensible compromise should be tailored to the type of specific data. In this study, we compromised with the time domain by using fPCA with a fixed number of eigendimensions while giving flexibility in the modelling of spatial dependence. This served the purpose of breaking off from the separability assumption while, at the same time, favouring interpretation and a simple model structure. The fact that the fPCs are treated as known for posterior inference might affect posterior uncertainty. One possible solution to improve coverage is to employ simultaneous credible bounds. These are a finite collection of point-wise intervals, scaled to achieve a specified coverage probability. Existing approaches include those of Besag et al.[64]; Krivobokova et al.[65] and Crainiceanu et al.[66]

By means of a simulation study and the analysis of fMRI and EEG data, we demonstrate that PCl-fPCA is effective in recovering the underlying smooth curves and it produces a valuable exploration of the spatio-temporal dependence in brain time series. The next step in our approach is the extension to the modelling of multiple subjects' recordings. There are different challenges to consider in the analysis of groups such as the natural inter-individual variability in brain functioning and the dimensionality of the data. We intend to expand our method to replicated data and multiple subjects experiments in our future research. Exploring inter-individual patterns of functional connectivity and their uncertainty can help answer important questions not only in the study of brain processes but also in the characterisation, early diagnosis and prognosis of brain diseases.

ACKNOWLEDGMENTS

fMRI data were kindly provided by IRCCS Santa Maria Nascente, Don Gnocchi foundation of Milan (Italy). EEG data were kindly provided by the Henri Begleiter Neurodynamics Laboratory, Department of Psychiatry and Behavioral Sciences, State University of New York (US). We thank the associate editor and reviewers for their useful comments and advice.

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

SUPPORTING INFORMATION

The following supporting information is available as part of the online article:
Supplementary material pdf file:
WebA - Sensitivity Analysis: In the sensitivity analysis we tested how changing the prior expected number of clusters and cluster size impacted on our findings in the application on fMRI data of Section 4 .
WebB - Additional Figures: Additional figures not shown in the paper that further clarify the features of our model, simulation study and application.
WebC - Simulation study setting: Additional information on the setting of the simulation study in Section 3.1.
WebD - Model checks: Additional information on the MCMC checks.

DATA AVAILABILITY

Please note that, upon publication, software in the form of R code will be available from an online repository together with the sample simulated data. EEG data are available at UCI Machine Learning Repository.[67]

How to cite this article: N. Margaritella, V. Inácio, and R. King (2020), Parameter clustering in Bayesian functional PCA of neuroscientific data, *Stat. Med.*, *2020;00:–*.

APPENDIX

A POSTERIOR CONDITIONAL DISTRIBUTIONS

In this section we present the posterior conditional distributions for the parameters of our model (Section 2.2).

$$\xi_{ik}|y_{it}, c_{i,k}, \mu_{jk}, s_{jk}, \tau \sim \mathrm{N}\left(\frac{\tau \sum_{t=1}^{T} y_{it}\phi_{tk} + s_{jk}\mu_{jk}}{\tau + s_{jk}}, \quad \frac{1}{\tau + s_{jk}}\right),$$

$$\tau|\tilde{\mathbf{y}}_1, ..., \tilde{\mathbf{y}}_n, a', b' \sim \Gamma\left(\frac{Tn}{2} + a', \quad \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\tilde{y}_{it} - \sum_{k=1}^{K}\xi_{ik}\phi_{kt}\right)^2}{2} + b'\right),$$

$$\mu_{jk}|\boldsymbol{c}_k, \boldsymbol{\xi}_k, s_{jk}, v_k, r_k \sim \mathrm{N}\left(\frac{s_{jk}\sum_{i:c_{ik}=j}^{n_{jk}}\xi_{ik} + v_k r_k}{n_{jk}s_{jk} + r_k}, \quad \frac{1}{n_{jk}s_{jk} + r_k}\right),$$

$$s_{jk}|\boldsymbol{c}_k, \boldsymbol{\xi}_k, \beta_k, z_k, \mu_{jk} \sim \Gamma\left(\frac{n_{jk}}{2} + z_k, \quad \frac{1}{2}\sum_{i:c_i=j}^{n_{jk}}(\xi_{ik} - \mu_{jk})^2 + \beta_k\right),$$

$$c_{ik}|\boldsymbol{p_k}, \boldsymbol{\xi_k}, \boldsymbol{\mu_k}, \boldsymbol{s_k}, \alpha_k \propto \sum_{j=1}^{J} p_{jk} s_{jk}^{1/2} \exp\left\{\frac{-s_{jk}}{2}(\xi_{ik} - \mu_{jk})^2\right\},$$

$$p_{1k} = \frac{p'_{1k}}{\sum_{j=1}^{J} p'_{jk}}; \quad p_{jk} = \frac{p'_{jk}\prod_{l<j}(1 - p_{lk})}{\sum_{j=1}^{J} p'_{jk}},$$

$$p'_{jk}|c_{ik}, \alpha_k \sim \mathrm{Beta}\left(n_{jk} + 1, \quad \alpha_k + \sum_{l=j+1}^{J} n_{lk}\right),$$

$$\alpha_k|\boldsymbol{p_k} \propto \alpha_k^J \exp\left\{\alpha_k \sum_{j=1}^{J}\log(1 - p'_{jk})\right\}; \quad \text{for } \mathcal{S}_{\alpha_k} = [0, Q_k],$$

where n_{jk} denote the fPC scores in the j^{th} cluster of the k^{th} eigendimension and $\mathcal{S}_{\alpha_k}$ the posterior support of α_k.

In our model we fixed $a' = b' = 10^{-3}$, $z_k = 1$, $v_k = 0$ and the upper-bound for the support of α_k takes into account the dimension-specific features of functional PCA as detailed in the paper, Section 2.2.

References

1. Ramsay J, Silverman BW. *Functional Data Analysis*. Springer Series in Statistics . 2005.
2. Ramsay J, Hooker G, Graves S. *Functional Data Analysis with R and MATLAB*. Springer Science & Business Media . 2009.
3. Crainiceanu CM, Goldsmith AJ. Bayesian functional data analysis using WinBUGS. *Journal of Statistical Software* 2010; 32(11).
4. Viviani R, Grön G, Spitzer M. Functional principal component analysis of fMRI data. *Human Brain Mapping* 2005; 24(2): 109–129.
5. Tian TS. Functional data analysis in brain imaging studies. *Frontiers in Psychology* 2010; 1: 35.
6. Hasenstab K, Scheffler A, Telesca D, et al. A multi-dimensional functional principal components analysis of EEG data. *Biometrics* 2017; 73(3): 999–1009.
7. Liu C, Ray S, Hooker G. Functional principal component analysis of spatially correlated data. *Statistics and Computing* 2017; 27(6): 1639–1654.
8. Wolfson O. Understanding the human brain via its spatio-temporal properties (vision paper). In: *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*ACM. ; 2018: 85–88.

9. Escobar MD, West M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 1995; 90(430): 577–588.

10. Neal RM. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 2000; 9(2): 249–265.

11. Rasmussen CE. The infinite Gaussian mixture model. In: *Advances in Neural Information Processing Systems*; 2000: 554–560.

12. James GM, Sugar CA. Clustering for sparsely sampled functional data. *Journal of the American Statistical Association* 2003; 98(462): 397–408.

13. Ray S, Mallick B. Functional clustering by Bayesian wavelet methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2006; 68(2): 305–332.

14. Zhou C, Wakefield J. A Bayesian mixture model for partitioning gene expression data. *Biometrics* 2006; 62(2): 515–525.

15. Dunson DB, Herring AH, Siega-Riz AM. Bayesian inference on changes in response densities over predictor clusters. *Journal of the American Statistical Association* 2008; 103(484): 1508–1517.

16. Bigelow JL, Dunson DB. Bayesian semiparametric joint models for functional predictors. *Journal of the American Statistical Association* 2009; 104(485): 26–36.

17. Rodríguez A, Dunson DB, Gelfand AE. Bayesian nonparametric functional data analysis through density estimation. *Biometrika* 2009; 96(1): 149–162.

18. Angelini C, De Canditiis D, Pensky M. Clustering time-course microarray data using functional Bayesian infinite mixture model. *Journal of Applied Statistics* 2012; 39(1): 129–149.

19. Scarpa B, Dunson DB. Enriched stick-breaking processes for functional data. *Journal of the American Statistical Association* 2014; 109(506): 647–660.

20. Xia Y, Chen Q, Shi L, et al. Tracking the dynamic functional connectivity structure of the human brain across the adult lifespan. *Human Brain Mapping* 2019; 40(3): 717–728.

21. Dong D, Duan M, Wang Y, et al. Reconfiguration of dynamic functional connectivity in sensory and perceptual system in schizophrenia. *Cerebral Cortex* 2019; 29(8): 3577–3589.

22. Fiorenzato E, Strafella AP, Kim J, et al. Dynamic functional connectivity changes associated with dementia in Parkinson's disease. *Brain* 2019; 142(9): 2860–2872.

23. Hutchison RM, Womelsdorf T, Allen EA, et al. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* 2013; 80: 360–378.

24. Wang JL, Chiou JM, Mueller HG. Review of functional data analysis. *Annual Review of Statistics and its Application* 2016; 3: 257–295.

25. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2016; 374(2065): 20150202.

26. Suarez AJ, Ghosal S, others . Bayesian estimation of principal components for functional data. *Bayesian Analysis* 2017; 12(2): 311–333.

27. Sørensen H, Goldsmith J, Sangalli LM. An introduction with medical applications to functional data analysis. *Statistics in Medicine* 2013; 32(30): 5222–5240.

28. Dunson DB. Nonparametric Bayes applications to biostatistics. In: *Bayesian Nonparametrics;*Cambridge University Press, Cambridge; 2010: 223–268.

29. Griffiths TL, Ghahramani Z. Infinite latent feature models and the indian buffet process. tech. rep., University College London, Gatsby Computational Neuroscience Unit; 2005.

30. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* 2006; 1(3): 515–534.

31. Richardson S, Green PJ. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1997; 59(4): 731–792.

32. Sethuraman J. A constructive definition of Dirichlet priors. *Statistica Sinica* 1994: 639–650.

33. Medvedovic M, Sivaganesan S. Bayesian infinite mixture model based clustering of gene expression profiles. *Bioinformatics* 2002; 18(9): 1194–1206.

34. Medvedovic M, Yeung KY, Bumgarner RE. Bayesian mixture model based clustering of replicated microarray data. *Bioinformatics* 2004; 20(8): 1222–1232.

35. De Iorio M, Gallot N, Valcarcel B, Wedderburn L. A Bayesian semiparametric Markov regression model for juvenile dermatomyositis. *Statistics in Medicine* 2018; 37(10): 1711–1731.

36. McDowell IC, Manandhar D, Vockley CM, Schmid AK, Reddy TE, Engelhardt BE. Clustering gene expression time series data using an infinite Gaussian process mixture model. *PLoS Computational Biology* 2018; 14(1): e1005896.

37. Müller P, Quintana FA, Jara A, Hanson T. *Bayesian nonparametric data analysis*. Springer . 2015.

38. Wade S, Ghahramani Z, others . Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Analysis* 2018; 13(2): 559–626.

39. Ramsay JO, Graves S, Hooker G. *fda: Functional Data Analysis*. 2020. R package version 5.1.4.

40. Escobar MD. Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* 1994; 89(425): 268–277.

41. Jara A, García-Zattera MJ, Lesaffre E. A Dirichlet process mixture model for the analysis of correlated binary responses. *Computational Statistics & Data Analysis* 2007; 51(11): 5402–5415.

42. Plummer M. *rjags: Bayesian Graphical Models using MCMC*. 2019. R package version 4-10.

43. Gelman A, Stern HS, Carlin JB, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. Chapman and Hall/CRC. 3rd ed. 2013.

44. Gentle JE. *Computational Statistics*. Springer . 2009.

45. Rasheed HA, Aref R. Bayesian inference for parameter and reliability function of Inverse Rayleigh distribution under modified squared error loss function. *Australian Journal of Basic and Applied Sciences* 2016; 10(16): 241–248.

46. Hubert L, Arabie P. Comparing partitions. *Journal of Classification* 1985; 2(1): 193–218.

47. Chen JE, Rubinov M, Chang C. Methods and considerations for dynamic analysis of functional MR imaging data. *Neuroimaging Clinics* 2017; 27(4): 547–560.

48. Islam MK, Rastegarnia A, Yang Z. Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology* 2016; 46(4-5): 287–305.

49. Griffanti L, Salimi-Khorshidi G, Beckmann CF, et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* 2014; 95: 232–247.

50. Van Den Heuvel MP, Pol HEH. Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology* 2010; 20(8): 519–534.

51. Zhang Z, Zhang D, Wang Z, et al. Intrinsic neural linkage between primary visual area and default mode network in human brain: evidence from visual mental imagery. *Neuroscience* 2018; 379: 13–21.

52. Milner A. How do the two visual streams interact with each other?. *Experimental Brain Research* 2017; 235(5): 1297–1308.

53. Leech R, Braga R, Sharp DJ. Echoes of the brain within the posterior cingulate cortex. *Journal of Neuroscience* 2012; 32(1): 215–222.

54. Tomasi D, Volkow ND. Functional connectivity hubs in the human brain. *Neuroimage* 2011; 57(3): 908–917.

55. Zhang XL, Begleiter H, Porjesz B, Wang W, Litke A. Event related potentials during object recognition tasks. *Brain Research Bulletin* 1995; 38(6): 531–538.

56. Sur S, Sinha V. Event-related potential: An overview. *Industrial Psychiatry Journal* 2009; 18(1): 70.

57. Plöchl M, Ossandón JP, König P. Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in human neuroscience* 2012; 6: 278.

58. Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human learning and memory* 1980; 6(2): 174.

59. Durante D, Dunson DB. Bayesian inference and testing of group differences in brain networks. *Bayesian Analysis* 2018; 13(1): 29–58.

60. Warnick R, Guindani M, Erhardt E, Allen E, Calhoun V, Vannucci M. A bayesian approach for estimating dynamic functional network connectivity in fMRI data. *Journal of the American Statistical Association* 2018; 113(521): 134–151.

61. Petrone S, Guindani M, Gelfand AE. Hybrid Dirichlet mixture models for functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2009; 71(4): 755–782.

62. Dunson DB. Nonparametric Bayes local partition models for random effects. *Biometrika* 2009; 96(2): 249–262.

63. MacLehose RF, Dunson DB. Nonparametric Bayes kernel-based priors for functional data analysis. *Statistica Sinica* 2009: 611–629.

64. Besag J, Green P, Higdon D, Mengersen K. Bayesian computation and stochastic systems. *Statistical Science* 1995: 3–41.

65. Krivobokova T, Kneib T, Claeskens G. Simultaneous confidence bands for penalized spline estimators. *Journal of the American Statistical Association* 2010; 105(490): 852–863.

66. Crainiceanu CM, Ruppert D, Carroll RJ, Joshi A, Goodner B. Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics* 2007; 16(2): 265–288.

67. EEG Database Data Set. http://archive.ics.uci.edu/ml/datasets/EEG+Database; . Accessed: 2020-04-16.